

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77214 A2

- (51) International Patent Classification⁷: C12N 15/31, C07K 14/395, 14/47, C12Q 1/68
- (71) Applicant (for all designated States except US): THE JOHN HOPKINS UNIVERSITY [US/US]; Suite 903, 111 Market Place, Baltimore, MD 21201 (US).
- (21) International Application Number: PCT/US00/16223
- (72) Inventors; and
- (22) International Filing Date: 14 June 2000 (14.06.2000)
- (75) Inventors/Applicants (for US only): VELCULESCU, Victor [US/US]; Apartment C, 650 N. Calvert Street, Baltimore, MD 21202 (US). VOGELSTEIN, Bert [US/US]; 3700 Breton Way, Baltimore, MD 21208 (US). KINZLER, Kenneth [US/US]; 1403 Halkirk Way, Belair, MD 21015 (US).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/335,032 16 June 1999 (16.06.1999) US
- (74) Agents: KAGAN, Sarah, A. et al.; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/335,032 (CON)
Filed on 16 June 1999 (16.06.1999)
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,

[Continued on next page]

(54) Title: CHARACTERIZATION OF THE YEAST TRANSCRIPTOME

	1	2	3
TDH2/3			
	519	561	636
TEF1/2			
	396	229	379
NORF1			
	114	84	182
RNR4			
	9	111	7
RNR2			
	9	85	9
NORF5			
	0	49	0

(57) Abstract: Yeast genes which are differentially expressed during the cell cycle are described. They can be used to study, affect, and monitor the cell cycle of a eukaryotic cell. They can be used to obtain human homologs involved in cell cycle regulation. They can be used to identify antifungal agents and other classes of drugs. They can be formed into arrays on solid supports for interrogation of a cell's transcriptome under various conditions.

WO 00/77214 A2



IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

Published:

— Without international search report and to be republished upon receipt of that report.

- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CHARACTERIZATION OF THE YEAST TRANSCRIPTOME

This application is a continuation-in-part of co-pending application Serial No. 09/012,031 filed January 22, 1998, the disclosure of which is incorporated by reference herein. This invention was made with government support under CA57345 awarded by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD OF THE INVENTION

This invention is related to the characterization of the expressed genes of the yeast genome. More particularly, it is related to the identification and use of previously unrecognized genes.

BACKGROUND OF THE INVENTION

It is by now axiomatic that the phenotype of an organism is largely determined by the genes expressed within it. These expressed genes can be represented by a "transcriptome," conveying the identity of each expressed gene and its level of expression for a defined population of cells. Unlike the genome, which is essentially a static entity, the transcriptome can be modulated by both external and internal factors. The transcriptome thereby serves as a dynamic link between an organism's genome and its physical characteristics.

The transcriptome as defined above has not been characterized in any eukaryotic or prokaryotic organism, largely because of technological limitations. However, some general features of gene expression patterns were elucidated two decades ago through

RNA-DNA hybridization measurements (Bishop *et al.*, 1974; Hereford and Rosbash, 1977). In many organisms, it was thus found that at least three classes of transcripts could be identified, with either high, medium, or low levels of expression, and the number of transcripts per cell were estimated (Lewin, 1980). These data of course provided little information about the specific genes that were members of each class. Data on the expression levels of individual genes have accumulated as new genes were discovered. However, in only a few instances have the absolute levels of expression of particular genes been measured and compared to other genes in the same cell type.

Description of any cell's transcriptome would therefore provide new information useful for understanding numerous aspects of cell biology and biochemistry.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide isolated DNA molecules and methods of using such molecules to affect the cell cycle and identify candidate drugs.

These and other objects of the invention are achieved by providing the art with one or more of the embodiments described below.

According to one embodiment of the invention an isolated DNA molecule is provided. It comprises a coding sequence of a yeast gene selected from the group consisting of NORF genes comprising a SAGE tag as shown in SEQ ID NOS:67-811.

According to another embodiment of the invention a method of using NORF genes is provided. The method is for affecting the cell cycle of a cell. The method comprises the step of administering to a cell an isolated DNA molecule comprising a coding sequence of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M.

In yet another embodiment of the invention a method for screening candidate antifungal drugs is provided. The method comprises the steps of contacting a test substance with a yeast cell and monitoring expression of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M, wherein a test substance

which modifies the expression of the yeast gene is a candidate antifungal drug.

In still another embodiment of the invention a method for identifying human genes which are involved in cell cycle progression is provided. The method comprises the step of contacting human DNA with a probe which comprises at least 14 contiguous nucleotides of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M. A human DNA sequence which hybridizes to the probe is identified as a sequence of a candidate human gene which is involved in cell cycle progression.

The present invention provides probes which comprise at least 14 contiguous nucleotides of a NORF gene comprising a SAGE tag as shown in SEQ ID NOS:67-811.

The invention also provides an array of probes on a solid support. At least one probe in the array comprises at least 14 contiguous nucleotides of a NORF gene comprising a SAGE tag as shown in SEQ ID NOS:67-811.

Still another embodiment of the invention is a method of identifying a candidate drug as a member of a class of drugs having a characteristic effect on gene expression in a yeast cell. A yeast cell is contacted with a candidate drug. Expression of at least one NORF gene whose expression is affected by the class of drugs is monitored in the yeast cell. Detection of a difference in expression of the at least one NORF gene relative to expression in the absence of the candidate drug identifies the candidate drug as a member of the class of drugs.

These and other embodiments of the invention which will be apparent to those of skill in the art upon reading the detailed disclosure provided below, make available to the art hitherto unrecognized genes, and information about the expression of genes globally at the organismal level.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Schematic of SAGE Method and Genome Analysis. In applying SAGE to the analysis of yeast gene expression patterns, the 3' most NlaIII site was used to define a unique position in each transcript and to provide a site for ligation of

a linker with a BsmFI site. The type II enzyme BsmFI, which cleaves a defined distance from its non-palindromic recognition site, was then used to generate a 15bp SAGE tag (designated by the black arrows), which includes the NlaIII site. Automated sequencing of concatenated SAGE tags allowed the routine identification of about a thousand tags per 36-lane sequencing gel. Once sequenced, the abundance of each SAGE tag was calculated, and each tag was used to search the entire yeast genome to identify its corresponding gene. The lower panel shows a small region of Chromosome 15. Gray arrows indicate all potential SAGE tags (NlaIII sites) and black arrows indicate 3' most SAGE tags. The total number of tags observed for each potential tag is indicated above (+ strand) or below (- strand) the tag. As expected, the observed SAGE tags were associated with the 3' end of expressed genes.

Figure 2. Sampling of Yeast Gene Expression. Analysis of increasing amounts of ascertained tags reveals a plateau in the number of unique expressed genes. Triangles represent genes with known functions, squares represent genes predicted on the basis of sequence information, and circles represent total genes.

Figure 3. Virtual Rot. (A) Abundance Classes in the Yeast Transcriptome. The transcript abundance is plotted in reverse order on the abscissa, whereas the fraction of total transcripts with at least that abundance is plotted on the ordinate. The dotted lines identify the three components of the curve, 1, 2, and 3. This is analogous to a Rot curve derived from reassociation kinetics where the product of initial RNA concentration and time is plotted on the abscissa, and the percent of labeled cDNA that hybridizes to excess mRNA is plotted on the ordinate. (B) Comparison of Virtual Rot and Rot Components. Transitions and data from virtual Rot components were calculated from the data in Figure 3A, while data for Rot components were obtained from Hereford and Rosbash, 1977.

Figure 4. Chromosomal Expression Map for *S. cerevisiae*. Individual yeast genes were positioned on each chromosome according to their open reading frame (ORF) start coordinates. Abundance levels of tags corresponding to each gene are displayed on the vertical axis, with transcription from the + strand indicated above the abscissa and that from the - strand indicated below. Yellow bands at ends of the expanded chromosome represent telomeric regions that are undertranscribed (see text

for details).

Figure 5. Northern Blot Analysis of Representative Genes. TDH2/3, TEF1/2 and NORF1, are expressed relatively equally in all three states (lane 1, G2/M arrested; lane 2, S phase arrested; lane 3, log phase), while RNR4, RNR2, and NORF5 are highly expressed in S-phase arrested cells. The expression level observed by SAGE (number of tags) is noted below each lane and was highly correlated with quantitation of the Northern blot by PhosphorImager analysis ($r^2=0.97$).

TABLE LEGENDS

Table 1. Highly Expressed Genes. Tag represents the 10 bp SAGE tag adjacent to the NlaIII site; Gene represents the gene or genes corresponding to a particular tag (multiple genes that match unique tags are from related families, with an average identity of 93%); Locus and Description denote the locus name and functional description of each ORF, respectively; Copies/cell represents the abundance of each transcript in the SAGE library, assuming 15,000 total transcripts per cell and 60,633 ascertained transcripts.

Table 2. Expression of Putative Coding Sequences. Table column headings are the same as for Table 1.

Table 3. Expression of the most abundant NORF genes. SAGE Tag, Locus, and Copies/cell are the same as for Table 1; Chr and Tag Pos denote the chromosome and position of each tag; ORF Size denotes the size of the ORF corresponding to the indicated tag. In each case, the tag was located within or less than 250 bp 3' of the NORF.

Table 4. Expression of NORF genes. SAGE tag and Copies/cell are the same as for Table 1. Chr and Tag Pos denote the chromosome and position of each tag.

Table 5. Gene expression changes in different cell cycle phases. L denotes log phase; S denotes synthesis phase; G2/M denotes the mitotic phase. Tag Sequence represents the 10 bp SAGE tag adjacent to the NlaIII site; "ratio L to S" denotes the ratio of expression in log phase to expression in synthesis phase; "ratio S to G2/M" denotes the ratio of expression in synthesis phase to expression in G2/M phase; "ratio G2/M to L" denotes the ratio of expression in G2/M to log phase. #DIV/0! indicates

an increase in expression from 0; a value of 0 indicates a decrease in expression to 0; a value of 1 indicates no change; a value less than 1 indicates a decrease in expression; and a value greater than 1 indicates an increase in expression.

Table 6. Intergenic open reading frames that contain or are adjacent to observed SAGE tags. Copies/cell represents abundance of each mRNA transcript as in Table 1. Positive expression level indicates the tag is on the + strand of the chromosome; Negative expression level indicates the tag is on the - strand.

DETAILED DESCRIPTION

It is a discovery of the present invention that certain hitherto unknown genes (the NORFs) exist and are expressed in yeast. These genes, as well as other previously identified and previously postulated genes, can be used to study, monitor, and affect phases of cell cycle. The present invention identifies which genes are differentially expressed during the cell cycle. Differentially expressed genes can be used as markers of phases of the cell cycle. They can also be used to affect a change in the phase of the cell cycle. In addition, they can be used to screen for drugs which affect the cell cycle, by affecting expression of the genes. Human homologs of these eukaryotic genes are also presumed to exist, and can be identified using the yeast genes as probes or primers to identify the human homologs.

New genes termed NORFs (not previously assigned open reading frames) have been found. They are uniquely identified by their SAGE tags. In addition their entire nucleotide sequences are known and publicly available. In general, these were not previously identified as genes due to their small size. However, they have now been found to be expressed.

Differentially expressed yeast genes are those whose expression varies by a statistically significant difference (to greater than 95% confidence level) within different growth phases, particularly log phase, S phase, and G2/M. Preferably the difference is at least 10%, 25%, 50%, or 100%. In some cases, differentially expressed genes are not expressed at detectable levels in one or more cell cycle phases as determined by SAGE analysis. Genes which have been found to have differential expression characteristics include: NORF N^a 1, 2, 4, 5, 6, 17, 25, 27, TEF1/TEF2,

EN02, ADH1, ADH2, PGK1, CUP1A/CUP1B, PYK1, YKL056C, YMR116C, YEL033W, YOR182C, YCR013C, ribonucleotide reductase 2 and 4, and YJR085C. Differential expression can be detected by any means known in the art, such as hybridization to specific probes or immunological assays.

5 Isolated DNA molecules according to the invention contain less than a whole chromosome and can be genomic or cDNA, *i.e.*, lacking introns. Isolated DNA molecules can comprise a yeast gene or a coding sequence of a yeast gene involved in cell cycle progression, such as NORF genes which comprise SAGE tags as shown in SEQ ID NOS:67-811. Isolated DNA molecules which comprise yeast genes or
10 coding sequences of yeast genes comprising SAGE tags as shown in SEQ ID NOS:37-12,203 are also isolated DNA molecules of the invention. Isolated DNA molecules can also consist of a yeast gene or a coding sequence of a yeast gene which comprises a SAGE tag as shown in SEQ ID NOS:37-12,203 or 67-811.

Any technique for obtaining a DNA of known sequence may be used to obtain
15 isolated DNA molecules of the invention. Preferably they are isolated free of other cellular components such as membrane components, proteins, and lipids. They can be made by a cell and isolated, or synthesized using PCR or an automatic synthesizer. Methods for purifying and isolating DNA are routine and are known in the art.

To administer yeast genes to cells, any DNA delivery techniques known in the
20 art may be used, without limitation. These include liposomes, transfection, mating, transduction, transformation, viral infection, electroporation. Vectors for particular purposes and characteristics can be selected by the skilled artisan for their known properties. Cells which can be used as gene recipients are yeast and other fungi, mammalian cells, including humans, and bacterial cells.

25 Antifungal drugs can be identified using yeast cells as described herein. Expression of a differentially expressed NORF gene can be monitored by any means known in the art. When a test substance modifies the expression of such a differentially expressed gene, for example by increasing or decreasing its expression, it is a candidate drug for affecting the growth properties of fungi and may be useful
30 as an antifungal agent. Expression of more than one NORF gene can be monitored. For example, expression of 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 75, 100, 150, 250,

300, 350, 400, 450, or 500 or more NORF genes can be monitored in single or multiple assays.

Because differentially expressed genes are likely to be involved in cell cycle progression, it is likely that these genes are conserved among species. The differentially expressed NORF genes identified by the present invention can be used to identify homologs in humans and other mammals by contacting DNA from these mammals with a probe which comprises at least 10 contiguous nucleotides of a differentially expressed NORF gene. The DNA can be genomic or cDNA, as is known in the art. Means for identifying homologous genes among different species are well known in the art. Briefly, stringency of hybridization can be reduced so that imperfectly matching sequences hybridize. This can be in the context of *inter alia* Southern blots, Northern blots, colony hybridization or PCR. Any hybridization technique which is known in the art can be used. A DNA sequence which hybridizes to the probe is identified as a sequence of a candidate gene which is involved in cell cycle expression.

Probes according to the present invention are isolated DNA molecules which have at least 10, and preferably at least 12, 14, 16, 18, 20, or 25 contiguous nucleotides of a particular NORF gene or other differentially expressed gene. The probes may or may not be labeled. They may be used, for example, as primers for PCR assays, or for detection of gene expression for Southern or Northern blots or *in situ* hybridization. Preferably the probes are immobilized on a solid support. The solid support can be any surface to which a probe can be attached. Suitable solid supports include, but are not limited to, glass or plastic slides, tissue culture plates, microtiter wells, tubes, or particles such as beads, including but not limited to latex, polystyrene, or glass beads. Any method known in the art can be used to attach the a probe to the solid support, including use of covalent and non-covalent linkages, passive absorption, or pairs of binding moieties attached respectively to the probe and the solid support.

More preferably, probes are present on an array so that multiple probes can simultaneously hybridize to a single biological sample. The probes can be spotted onto the array or synthesized *in situ* on the array. See Lockhart et. al., *Nature*

Biotechnology, Vol. 14, December 1996, "Expression monitoring by hybridization to high-density oligonucleotide arrays." A single array contains at least one NORF probe, but can contain more than 100, 500 or even 1,000 different probes in discrete locations. If desired, one or more NORF probe(s) present on the array can be nucleotide sequences from a NORF gene which is differentially expressed during the cell cycle.

Genes identified by the present invention which are differentially expressed during the cell cycle can also be used to obtain gene expression profiles characteristic of the response of yeast genes of a yeast cell to a particular drug or class of drugs. Classes of drugs of particular interest for which gene expression profiles can be generated include those drugs which affect cell cycle or other cell processes, such as chemotherapeutic agents. If desired, gene expression profiles characteristic of more than one drug of a particular class can be generated and used to make a composite gene expression profile. For example, microtubule poison drugs such as vinblastin, taxol, vincristine, and taxotere can be used to generate gene expression profiles characteristic of microtubule poisons.

To generate a gene expression profile characteristic of a particular drug or class of drugs, a yeast cell is contacted with a particular drug or a member of a particular class of drugs. Expression of at least one yeast gene is monitored, either before and after contacting or in the contacted cell and in another yeast cell which has not been contacted with the drug. Genes which are monitored can be any yeast gene, including NORFS. Preferably, these genes are differentially expressed during the cell cycle. For example, yeast genes can be selected from genes comprising the SAGE tags shown in Tables 3, 4, 5, and 6 (SEQ ID NOS:67-12,203). If desired, genes such as NORF N^a 1, 2, 4, 5, 6, 17, 25, or 27, TEF1/TEF2, EN02, ADH1, ADH2, PGK1, CUP1A/CUP1B, PYK1, YKL056C, YMR116C, YEL033W, YOR182C, YCR013C, ribonucleotide reductase 2 and 4, and YJR085C, can be used for monitoring alterations in gene expression.

The expression of any number of these genes, such as 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250, 500, 1000, 2000, 3000, 4000, 5000, or 5,500 genes, can be measured. It is particularly convenient to monitor expression of the

differentially expressed genes using nucleic acids which are immobilized on a solid support or in an array, such as the gene arrays described above.

Many genes, particularly cell cycle genes, are likely to be conserved between yeast and mammals, including humans. Thus, gene expression profiles characteristic of a drug or class of drugs can be used to predict the effects of candidate drugs on human cells, by identifying the candidate drug as a member of a class of drugs whose characteristic gene expression profile is known. The candidate drugs can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity. The candidate drugs can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, and can be produced recombinantly or synthesized by chemical methods known in the art.

The effect of a candidate drug on expression of at least one gene whose expression is affected by the class of drugs is monitored. A gene expression profile obtained using the candidate drug which is similar to a gene expression profile for a particular drug or class of drugs identifies the candidate drug as a member of that class of drugs.

The effect of modifying particular substituents of a known drug or of a candidate drug can be similarly tested. Such methods are useful for determining whether alterations intended, for example, to increase solubility or absorption of a particular drug will have an unintended and possibly deleterious effect on genes which are differentially expressed during the cell cycle.

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

EXAMPLE

Summary

We have analyzed the set of genes expressed from the yeast genome, herein called the transcriptome, using serial analysis of gene expression (SAGE). Analysis

of 60,633 transcripts revealed 4,665 genes, with expression levels ranging from 0.3 to over 200 transcripts per cell. Of these genes, 1,981 had known functions, while 2,684 were previously uncharacterized. Integration of positional information with gene expression data allowed the generation of chromosomal expression maps, identifying physical regions of transcriptional activity, and identified genes that had not been predicted by sequence information alone. These studies provide insight into global patterns of gene expression in yeast and demonstrate the feasibility of genome-wide expression studies in eukaryotes.

Results

Characteristics and Rationale of SAGE Approach

Several methods have recently been described for the high throughput evaluation of gene expression (Nguyen *et al.*, 1995; Schena *et al.*, 1995; Velculescu *et al.*, 1995). We used SAGE (Serial Analysis of Gene Expression) because it can provide quantitative gene expression data without the prerequisite of a hybridization probe for each transcript. The SAGE technology is based on two basic principles (Figure 1). First, a short sequence tag (9-11 bp) contains sufficient information to uniquely identify a transcript, provided that it is derived from a defined location within that transcript. Second, many transcript tags can be concatenated into a single molecule and then sequenced, revealing the identity of multiple tags simultaneously. The expression pattern of any population of transcripts can be quantitatively evaluated by determining the abundance of individual tags and identifying the gene corresponding to each tag.

Genome-wide expression

In order to maximize representation of genes involved in normal growth and cell-cycle progression, SAGE libraries were generated from yeast cells in three states: log phase, S phase arrested and G2/M phase arrested. In total, SAGE tags corresponding to 60,633 total transcripts were identified (including 20,184 from log phase, 20,034 from S phase arrested, and 20,415 from G2/M phase arrested cells). Of these tags, 56,291

tags (93%) precisely matched the yeast genome, 88 tags matched the mitochondrial genome, and 91 tags matched the 2 micron plasmid.

The number of SAGE tags required to define a yeast transcriptome depends on the confidence level desired for detecting low abundance mRNA molecules. Assuming the previously derived estimate of 15,000 mRNA molecules per cell (Hereford and Rosbash, 1977), 20,000 tags would represent a 1.3 fold coverage even for mRNA molecules present at a single copy per cell, and would provide a 72% probability of detecting such transcripts (as determined by Monte Carlo simulations). Analysis of 20,184 tags from log phase cells identified 3,298 unique genes. As an independent confirmation of mRNA copy number per cell, we compared the expression level of *SUP44/RPS4*, one of the few genes whose absolute mRNA levels have been reliably determined by quantitative hybridization experiments (Iyer and Struhl, 1996), with expression levels determined by SAGE. *SUP44/RPS4* was measured by hybridization at 75 +/- 10 copies/cell (Iyer and Struhl, 1996), in good accord with the SAGE data of 63 copies/cell, suggesting that the estimate of 15,000 mRNA molecules per cell was reasonably accurate. Analysis of SAGE tags from S phase arrested and G2/M phase arrested cells revealed similar expression levels for this gene (range 52 to 55 copies/cell), as well as for the vast majority of expressed genes. As less than 1% of the genes were expressed at dramatically different levels among these three states (see below), SAGE tags obtained from all libraries were combined and used to analyze global patterns of gene expression.

Analysis of ascertained tags at increasing increments revealed that the number of unique transcripts plateaued at ~60,000 tags (Figure 2). This suggested that generation of further SAGE tags would yield few additional genes, consistent with the fact that sixty thousand transcripts represented a four-fold redundancy for genes expressed as low as 1 transcript per cell. Likewise, Monte Carlo simulations indicated that analysis of 60,000 tags would identify at least one tag for a given transcript 97% of the time if its expression level was one copy per cell.

The 56,291 tags that precisely matched the yeast genome represented 4,665 different genes. This number is in agreement with the estimate of 3,000 to 4,000 expressed genes obtained by RNA-DNA reassociation kinetics (Hereford and

Rosbash, 1977). These expressed genes included 85% of the genes with characterized functions (1,981 of 2,340), and 76% of the total genes predicted from analysis of the yeast genome (4,665 of 6,121). These numbers are consistent with a relatively complete sampling of the yeast transcriptome given the limited number of physiological states examined and the large number of genes predicted solely on the basis of genomic sequence analysis.

The transcript expression per gene was observed to vary from 0.3 to over 200 copies per cell. Analysis of the distribution of gene expression levels revealed several abundance classes that were similar to those observed in previous studies using reassociation kinetics. A "virtual Rot" of the genes observed by SAGE (Figure 3A) identified three main components of the transcriptome with abundances ranging over three orders of magnitude. A Rot curve derived from RNA-cDNA reassociation kinetics also contained three main components distributed over a similar range of abundances (Hereford and Rosbash, 1977). Although the kinetics of reassociation of a particular class of RNA and cDNA may be affected by numerous experimental variables, there were striking similarities between Rot and virtual Rot analyses (Figure 3B). Because Rot analysis may not detect all transcripts of low abundance (Lewin, 1980), it is not surprising that SAGE revealed both a larger total number of expressed genes and a higher fraction of the transcriptome belonging to the low abundance transcript class.

Integration of Expression Information with the Genomic Map

The SAGE expression data could be integrated with existing positional information to generate chromosomal expression maps (Figure 4). These maps were generated using the sequence of the yeast genome and the position coordinates of ORFs obtained from the Stanford Yeast Genome Database. Although there were a few genes that were noted to be physically proximal and have similarly high levels of expression, there did not appear to be any clusters of particularly high or low expression on any chromosome. Genes like histones H3 and H4, which are known to have coregulated divergent promoters and are immediately adjacent on chromosome 14 (Smith and Murray, 1983), had very similar expression levels (5 and 6 copies per cell,

respectively). The distribution of transcripts among the chromosomes suggested that overall transcription was evenly dispersed, with total transcript levels being roughly linearly related to chromosome size ($r^2 = 0.85$, data not shown). However, regions within 10 kb of telomeres appeared to be uniformly undertranscribed, containing on average 3.2 tags per gene as compared with 12.4 tags per gene for non-telomeric regions (Figure 4). This is consistent with the previously described observations of "telomeric silencing" in yeast (Gottschling *et al.*, 1990). Recent studies have reported telomeric position effects as far as 4 kb from telomere ends (Renauld *et al.*, 1993).

Gene Expression Patterns

Table 1 lists the 30 most highly expressed genes, all of which are expressed at greater than 60 mRNA copies per cell. As expected, these genes mostly correspond to well characterized enzymes involved in energy metabolism and protein synthesis and were expressed at similar levels in all three growth states (Examples in Figure 5). Some of these genes, including *ENO2* (McAlister and Holland, 1982), *PDC1* (Schmitt *et al.*, 1983), *PGK1* (Chambers *et al.*, 1989), *PYK1* (Nishizawa *et al.*, 1989), and *ADH1* (Denis *et al.*, 1983), are known to be dramatically induced in the glucose-rich growth conditions used in this study. In contrast, glucose repressible genes such as the *GAL1/GAL7/GAL10* cluster (St John and Davis, 1979), and *GAL3* (Bajwa *et al.*, 1988) were observed to be expressed at very low levels (0.3 or fewer copies per cell).

As expected for the yeast strain used in this study, mating type α specific genes, such as the α factor genes (*MFA1*, *MFA2*) (Michaelis and Herskowitz, 1988), and alpha factor receptor (*STE2*) (Burkholder and Hartwell, 1985) were all observed to be expressed at significant levels (range 2 to 10 copies per cell), while mating type α specific genes (*MFA1*, *MFA2*, *STE3*) (Hagen *et al.*, 1986; Kurjan and Herskowitz, 1982; Singh *et al.*, 1983) were observed to be expressed at very low levels (<0.3 copies/cell).

Three of the highly expressed genes in Table 1 had not been previously characterized. One contained an ORF with predicted ribosomal function, previously identified only by genomic sequence analysis. Analyses of all SAGE data suggested that there were 2,684 such genes corresponding to uncharacterized ORFs which were

transcribed at detectable levels. The 30 most abundant of these transcripts were observed more than 30 times, corresponding to at least 8 transcripts per cell (Table 2).

The other two highly expressed uncharacterized genes corresponded to ORFs not predicted by analysis of the yeast genome sequence (NORF = Nonannotated ORF).

Analyses of SAGE data suggested that there were at least 160 *NORF* genes transcribed at detectable levels. The 30 most abundant of these transcripts were observed at least 9 times (Table 3 and examples in Figure 5).

Interestingly, one of the *NORF* genes (*NORF5*) was only expressed in S phase arrested cells and corresponded to the transcript whose abundance varied the most in the three states analyzed (> 49 fold, Figure 5). Comparison of S phase arrested cells to the other states also identified greater than 9 fold elevation of the *RNR2* and *RNR4* transcripts (Figure 5). Induction of these ribonucleoside reductase genes is likely to be due to the hydroxyurea treatment used to arrest cells in S phase (Elledge and Davis, 1989). Likewise, comparison of G2/M arrested cells identified elevation of *RBL2* and dynein light chain, both microtubule associated proteins (Archer *et al.*, 1995; Dick *et al.*, 1996). As with the RNR inductions, these elevated levels seem likely to be related to the nocodazole treatment used to arrest cells in the G2/M phase. While there were many relatively small differences between the states (for example, *NORF1*, Figure 5), overall comparison of the three states revealed surprisingly few dramatic differences; there were only 29 transcripts whose abundance varied more than 10 fold among the three different states analyzed (Tables 4 and 5).

A comprehensive analysis for NORF genes was performed using the SAGE data. Yeast genome intergenic regions were defined as regions outside annotated ORFs or the 500bp region downstream of annotated ORFs (yeast genome sequence and tables of annotated ORFs were obtained from SGD at <http://genome-www.stanford.edu/Saccharomyces/>). Based on sequence analysis a total of 9524 putative ORFs of 25-99 amino acids were present in the intergenic regions; 510 of these ORFs contain or are adjacent to observed SAGE tags (Table 6). Of the 60,633 SAGE tags analyzed, there were 302 unique SAGE tags either within or adjacent to intergenic ORFs (100bp upstream or 500bp downstream of the ORF) (Table 6). Note that in some cases, more than one NORF contains or is adjacent to the

SAGE tag. These tags matched the genome uniquely, were in the correct orientation, and were expressed at levels greater than 0.3 transcript copies per cell.

The expression level for each NORF shown in Table 6 corresponds to the number of mRNA transcript copies per cell. If the expression level is positive it means that the tag is on the + strand of the chromosome; if negative, the tag is on the – strand of the chromosome.

Discussion

Analysis of a yeast transcriptome affords a unique view of the RNA components defining cellular life. Comparison of gene expression patterns from altered physiologic states can provide insight into genes that are important in a variety of processes. Comparison of transcriptomes from a variety of physiologic states should provide a minimum set of genes whose expression is required for normal vegetative growth, and another set composed of genes that will be expressed only in response to specific environmental stimuli, or during specialized processes. For example, recent work has defined a minimal set of 250 genes required for prokaryotic cellular life (Mushegian and Koonin, 1996). Examination of the yeast genome readily identified homologous genes for 196 of these, over 90% of which were observed to be expressed in the SAGE analysis. Detailed analyses of yeast transcriptomes, as well as transcriptomes from other organisms, should ultimately allow the generation of a minimal set of genes required for eukaryotic life.

Like other genome-wide analyses, SAGE analysis of yeast transcriptomes has several potential limitations. First, a small number of transcripts would be expected to lack an NlaIII site and therefore would not be detected by our analysis. Second, our analysis was limited to transcripts found at least as frequently as 0.3 copies per cell. Transcripts expressed in only a minute fraction of the cell cycle, or transcripts expressed in only a fraction of the cell population, would not be reliably detected by our analysis. Finally, mRNA sequence data are practically unavailable for yeast, and consequently, some SAGE tags cannot be unambiguously matched to corresponding genes. Tags which were derived from overlapping genes, or genes which have unusually long 3' untranslated regions may be misassigned. Increased availability of

3' UTR sequences in yeast mRNA molecules should help to resolve the ambiguities.

Despite these potential limitations, it is clear that the analyses described here furnish both global and local pictures of gene expression, precisely defined at the nucleotide level. These data, like the sequence of the yeast genome itself, provide simple, basic information integral to the interpretation of many experiments in the future. The availability of mRNA sequence information from EST sequencing as well as various genome projects, will soon allow definition of transcriptomes from a variety of organisms, including human. The data recorded here suggest that a reasonably complete picture of a human cell transcriptome will require only about 10 - 20 fold more tags than evaluated here, a number well within the practical realm achievable with a small number of automated sequencers. The analysis of global expression patterns in higher eukaryotes is expected, in general, to be similar to those reported here for *S. cerevisiae*. However, the analysis of the transcriptome in different cells and from different individuals should yield a wealth of information regarding gene function in normal, developmental, and disease states.

Experimental Procedures

Yeast cell culture

The source of transcripts for all experiments was *S. cerevisiae* strain YPH499 (*MATa ura3-52 lys2-801 ade2-101 leu2-Δ1 his3-Δ200 trp1-Δ63*) (Sikorski and Hieter, 1989). Logarithmically growing cells were obtained by growing yeast cells to early log phase (3×10^6 cells/ml) in YPD (Rose *et al.*, 1990) rich medium (YPD supplemented with 6 mM uracil, 4.8 mM adenine and 24 mM tryptophan) at 30°C. For arrest in the G1/S phase of the cell cycle, hydroxyurea (0.1 M) was added to early log phase cells, and the culture was incubated an additional 3.5 hours at 30°C. For arrest in the G2/M phase of the cell cycle, nocodazole (15 μg/ml) was added to early log phase cells and the culture was incubated for an additional 100 minutes at 30°C. Harvested cells were washed once with water prior to freezing at -70°C. The growth states of the harvested cells were confirmed by microscopic and flow cytometric analyses (Basrai *et al.*, 1996).

SAGE protocol

The SAGE method was performed as previously described (Velculescu *et al.*, 1995; Kinzler *et al.*, U.S. Patents 5,866,330 and 5,695,937), with exceptions noted below. PolyA RNA was converted to double-stranded cDNA with a BRL synthesis kit using the manufacturer's protocol except for the inclusion of primer biotin-5'-T₁₈-3'. The cDNA was cleaved with NlaIII (Anchoring Enzyme). As NlaIII sites were observed to occur once every 309 base pairs in three arbitrarily chosen yeast chromosomes (1, 5, 10), 95% of yeast transcripts were predicted to be detectable with a NlaIII-based SAGE approach. After capture of the 3' cDNA fragments on streptavidin coated magnetic beads (Dynal), the bound cDNA was divided into two pools, and one of the following linkers containing recognition sites for BsmFI was ligated to each pool: Linker 1, 5'-TTTGGATTGCTGGTGCAGTACAAGCTTAATAGGGACATG-3' (SEQ ID NO:1); 5'-TCCCTATTAAGCCTAGTTGTACTGCACCAGCAAATCC [amino mod. C7]-3' (SEQ ID NO:2); Linker 2, 5'-TTTCTGCTCGAATTCAAGCTTCTAACGATGTACGGGGACATG-3' (SEQ ID NO:3), 5'-TCCCCGTACATCGTTAGAAGCTTGAATTCGAGCAG [amino mod. C7]-3' (SEQ ID NO:4).

As BsmFI (Tagging Enzyme) cleaves 14 bp away from its recognition site, and the NlaIII site overlaps the BsmFI site by 1 bp, a 15 bp SAGE tag was released with BsmFI. SAGE tag overhangs were filled-in with Klenow, and tags from the two pools were combined and ligated to each other. The ligation product was diluted and then amplified with PCR for 28 cycles with 5'-GGATTGCTGGTGCAGTACA-3' (SEQ ID NO:5) and 5'-CTGCTCGAATTCAAGCTTCT-3' (SEQ ID NO:6), as primers. The PCR product was analyzed by polyacrylamide gel electrophoresis (PAGE), and the PCR product containing two tags ligated tail to tail (ditag) was excised. The PCR product was then cleaved with NlaIII, and the band containing the ditags was excised and self-ligated. After ligation, the concatenated products were separated by PAGE and products between 500 bp and 2 kb were excised. These products were cloned into the SphI site of pZero (Invitrogen). Colonies were screened for inserts by PCR with M13 forward and M13 reverse sequences located outside the cloning site as primers.

PCR products from selected clones were sequenced with the TaqFS DyePrimer kits (Perkin Elmer) and analyzed using a 377 ABI automated sequencer (Perkin Elmer), following the manufacturer's protocol. Each successful sequencing reaction identified an average of 26 tags; given a 90% sequencing reaction success rate, this corresponded to an average of about 850 tags per sequencing gel.

SAGE data analysis

Sequence files were analyzed by means of the SAGE program group (Velculescu *et al.*, 1995), which identifies the anchoring enzyme site with the proper spacing and extracts the two intervening tags and records them in a database. The 68,691 tags obtained contained 62,965 tags from unique ditags and 5,726 tags from repeated ditags. The latter were counted only once to eliminate potential PCR bias of the quantitation, as described (Velculescu *et al.*, 1995). Of 62,965 tags, 2,332 tags corresponded to linker sequences, and were excluded from further analysis. Of the remaining tags, 4,342 tags could not be assigned, and were likely due to sequencing errors (in the tags or in the yeast genomic sequence). If all of these were due to tag sequencing errors, this corresponds to a sequencing error rate of about 0.7% per base pair (for a 10bp tag), not far from what we would have expected under our automated sequencing conditions. However, some unassigned tags had a much higher than expected frequency of A's as the last five base pairs of the tag (5 of the 52 most abundant unassigned tags), suggesting that these tags were derived from transcripts containing anchoring enzyme sites within several base pairs from their polyA tails. Given the frequency of NlaIII sites in the genome (one in 309 base pairs), approximately 3% of transcripts were predicted to contain NlaIII sites within 10 bp of their polyA tails.

As very sparse data are available for yeast mRNA sequences and efforts to date have not been able to identify a highly conserved polyadenylation signal (Irniger and Braus, 1994; Zaret and Sherman, 1982), we used 14 bp of SAGE tags (i.e. the NlaIII site plus the adjacent 10 bp) to search the yeast genome directly (yeast genome sequence obtained from the Stanford yeast genome ftp site (genome-ftp.stanford.edu) on August 7, 1996). Because only coding regions are annotated in the yeast genome,

and SAGE tags can be derived from 3' untranslated regions of genes, a SAGE tag was considered to correspond to a particular gene if it matched the ORF or the region 500 bp 3' of the ORF (locus names, gene names and ORF chromosomal coordinates were obtained from Stanford yeast genome ftp site, and ORF descriptions were obtained from MIPS www site (<http://www.mips.biochem.mpg.de/>) on August 14, 1996). ORFs were considered genes with known functions if they were associated with a three letter gene name, while ORFs without such designations were considered uncharacterized.

As expected, SAGE tags matched transcribed portions of the genome in a highly non-random fashion, with 88% matching ORFs or their adjacent 3' regions in the correct orientation (chi-squared P value $<10^{-30}$). In instances when more than one tag matched a particular ORF in the correct orientation, the abundance was calculated to be the sum of the matched tags. Tags that matched ORFs in the incorrect orientation were not used in abundance calculations. In instances when a tag matched more than one region of the genome (for example an ORF and non-ORF region) only the matched ORF was considered. In some cases the 15th base of the tag could also be used to resolve ambiguities.

For the identification of NORF genes, only tags were considered that matched portions of the genome that were further than 500 bp 3' of a previously identified ORF and were observed at least two times in the SAGE libraries.

Table 1. Highly expressed genes

Tag	Seq. ID No.	Gene	Locus	Copies/cell	Description
GGGTTAACG	7	TDH2/TDH3	YIR006C/YOR112C	425	glyoxaldehyde-3-phosphate dehydrogenase 2 & 3
AGACAACCTG	8	TEF1/TEF2	YPR009W/YBR118W	248	cytosolic elongation factor eEF-1 alpha-A chain
TACCACTCCT	8	ENO2	YHR174W	229	2-phosphoglycerate dehydratase
GGTTTGCGT	10	RPLA1, A2, A3, 10E	YOL081C/YOL038W/YOL130W/YLR040W	207	acidic ribosomal protein a1 / P2, beta / L44prime / L10
TTGGCAGCT	11	PDC1	YLR044C	207	pyruvate decarboxylase isozyme 1
GGTGAAMACG	12	ADH1, ADH2	YOL086C/YHR030C	182	sialonid dehydrogenase I / II
ATGCGCCGCT	13	GPM1	YOL123C	168	phosphoglycerate mutase
GGTGTAAAGA	14	FBA1	YOL090C	166	fructose-bisphosphate aldolase II
TTAGTTTCTA	15	RPL47A	YOL104C	143	ribosomal protein
TCTCTACTGG	16	PGK1	YCR012W	139	phosphoglycerate kinase
GGTTTGGTT	17	RPL44	YDR382W	138	acidic ribosomal protein L45
GGTCCAGCTT	18	SSM1A / SSM1B	YPL220W / YGL135W	128	ribosomal protein
AATCCAGTTG	19	RPL5A / RPL5B	YLD18W / YFR031AC	102	ribosomal protein
TTGGTTGACT	20		NOR1	84	nontranslated ORF
AACAGACACG	21	RPL16A / RPL16B	YFR102C / YGR055C	83	ribosomal protein
CTGCTCTGGG	22	CLP1A / CLP1B	YHR033C / YHR055C	75	metallothionein
GCAATACTAC	23		YOR235W / YMR230W	73	ribosomal protein S10 / similarly to ribosomal protein S10
GCTCTCCGCC	24		NOR2	73	nontranslated ORF
AAAGACAGAG	25	RPS31A	YOR027C	72	ribosomal protein
TGTGTTGGTG	26	RPL2A / RPL2B	YBR031W / YOR012W	70	ribosomal protein
CCAAGGGTAT	27	RPS28A	YGR118W	69	ribosomal protein
TCTCCAGAAAG	28	RPL35B	YOR500C	69	ribosomal protein
GTTTTCCTTT	29	PYK1	YAL039W	68	pyruvate kinase
ATCACTGGTG	30	RPL6A / RPL6B	YGL147C / YNL057W	68	ribosomal protein L9
ATGAAGGTTT	31	RPL27A	YHR010W	68	ribosomal protein L27
GTAGAGGCGG	32	RPS21	YOL040C	67	ribosomal protein
GGTACTGATG	33	RPL43A	YDL075W	67	ribosomal protein
CCAGATTGTG	34	NAB1A / NAB1B	YGR214W / YLR048W	67	40S ribosomal protein p40 homolog A
GTGCGCTCCA	35	URP1A	YBR181W	62	ribosomal protein L21
CAAAACCCAA	36	RPS18EB	YML026C	60	ribosomal protein S18

Table 2. Putative coding sequences

SAGE Tag	Seq. ID No.	Locus	Copies/Cell	Description
TTGAAGTACC	37	YKL056C	58	strong similarity to human IgE-dependent histamine-releasing factor (21K tumor protein)
TTCGGGTAC	38	YDR278C	56	strong similarity to Hordeum vulgare b1t101 protein
CCAGATATGA	39	YIL093C	41	hypothetical protein
TTTAAATGG	40	YMR116C	38	similarity to N.crassa CPC2 protein
GGTGTGCTG	41	YBR078W	34	strong similarity to sporulation specific Sps2p
TACTCTTCGC	42	YEL033W	33	hypothetical protein
TGTAATTAA	43	YOR182C	26	homology to human ubiquitin-like protein/ribosomal protein S30
GGAGATCTTG	44	YCR013C	24	weak similarity to M.lepra B1496_F1_41 protein
TCAAGAAAGT	45	YER056AC	20	strong similarity to ribosomal protein L34
AAAACTTGG	46	YIL051C	18	strong similarity to YER057c
AAGTTGAACA	47	YPR043W	17	ribosomal protein L37
GGGTGCGGGT	48	YDR032C	16	strong similarity to YCR004c and S.pombe obr1
TGACTCTTGG	49	YLR390W	14	hypothetical protein
GGTCAATGGC	50	YJR105W	11	hypothetical protein
TAAGAATTCT	51	YJL158C	11	member of the Pir1p/Hsp150p/Pir3p family
TCAATTATGT	52	YDR033W	11	strong similarity to putative heat shock protein YRO2
ACGGCCCAAGA	53	YBR162C	10	similarity to YJL171p
TTGGGCTAGT	54	YJL171C	10	similarity to YBR162c
CCTTCAGGT	55	YJR085C	10	hypothetical protein
CCTCTCTGT	56	YOR310C	10	homology to SIK1 protein
CCCAAAACTT	57	YEL018W	9	weak similarity to Rad50p
AACAAGTACT	58	YGL037C	9	similarity to E.coli hypothetical 23K protein
AACAATAAAA	59	YER072W	8	similarity to YFL004w
CAAAAAGACCG	60	YML056C	8	homology to human IMP dehydrogenase I
GGTTTTTGAT	61	YOR182C	8	homology to human ubiquitin-like protein/ribosomal protein S30
CAATCCATT	62	YBR106W	8	hypothetical protein
TTTTGGGCTCT	63	YMR318C	8	putative alcohol-dehydrogenase
AACTGTCCAT	64	YDR429C	8	similarity to nuclear RNA binding proteins
CCAAGGTAA	65	YAR002AC	8	strong similarity to YGL002w
GGTTTTTGAA	66	YOR273C	8	putative resistance protein

Table 3. NORF genes

SAGE Tag	Seq. ID No.	Locus	Copies/Cell	Chr	Tag Pos	ORF Size (bp)
TTCGTTCACT	67	NORF1	94	4	1489450	198
GCTCTCCCC	68	NORF2	73	16	75633	243
TGTACGCATT	69	NORF3	16	15	301251	189
TTTTATTATC	70	NORF4	15	6	223182	177
CTTCTCTTTT	71	NORF5	12	13	158973	204
TTTCTATATA	72	NORF6	11	13	511754	252
TCTAGTCGCC	73	NORF7	10	12	669659	192
ATCGTTTAT	74	NORF8	8	15	877140	174
GGCCAATGGT	75	NORF9	8	4	1202289	267
ACCCTGTCAT	76	NORF10	7	2	418633	255
AAAAGATCAT	77	NORF11	7	4	1489453	87
CAGAAAAATGG	78	NORF12	6	8	115655	279
TGACATTCTT	79	NORF13	6	16	883669	183
TAGACATCTA	80	NORF14	6	2	491117	141
TGCCCTGGCC	81	NORF15	5	5	166452	216
GGTTTGGCG	82	NORF16	4	3	24169	291
CCATACAGGT	83	NORF17	4	12	673851	114
CCAAATCAAA	84	NORF18	3	4	229494	258
AAGCGGTACT	85	NORF19	3	9	47889	399
AACGCTTTTC	86	NORF20	3	2	351456	198
GAGGATAGAG	87	NORF21	3	2	358201	240
CAATGAACCG	88	NORF22	3	16	75541	243
TCCTTATATA	89	NORF23	3	1	73363	90
CGCCTCCAGT	90	NORF24	3	7	485774	108
TACGTAAGTT	91	NORF25	3	10	156139	81
GAATTAACT	92	NORF26	3	15	254749	93
GCGCCTCCAA	93	NORF27	2	5	42622	222
CAATGGCCCA	94	NORF28	2	13	511751	78
TTGAGGAACG	95	NORF29	2	3	154681	264
GCTAAGAACC	96	NORF30	2	4	302607	204

Additional NORFs

SAGE Tag	Seq. ID No.	Chr	Tag Pos	Copies/cell
GGCGCAATTT	97	4	1108395	2
TAAGTGATGA	98	7	593382	2
TTGTTGAATT	99	10	608373	2
GAAGCAGTAA	100	3	155607	2
ACATATGTTA	101	4	916112	2
CCCTACACGG	102	6	223289	2
GTAATTGGAC	103	10	392099	2
ATCAGACAAA	104	14	687272	2
TTATGAAAGA	105	15	81263	2
ATTCGTTCTA	106	15	841970	2
AGCAGGAGTT	107	16	188350	2
TTCTATTAGG	108	2	418749	2
TGGATTTTCAG	109	4	1224930	2
CAGATATAAT	110	5	52488	2
CTGTTTTGGG	111	11	374761	2
CATTTTTAGT	112	11	508212	2
TTGAAAAGAT	113	13	104160	2
TAAGCCCATC	114	13	251273	2
AGCGTCCTCA	115	15	832420	2
TTAGTTAAT	116	2	477623	2
ATGGTAGCCA	117	3	56961	2
AATTAGACTA	118	3	162589	2
AGTGACTCTT	119	4	1490879	2
GGACTATAAG	120	5	251266	2
ACTTTTTTCAG	121	10	159213	2
GTCATATAGT	122	13	158765	2
CAACAAAGTG	123	13	171166	2
GTGGGAAAGG	124	13	804600	2
TACTTTATAT	125	16	366449	2
AATACCAGCG	126	3	175540	1
GCCTTGATA	127	4	372624	1
GGTACATTCA	128	5	67152	1
GATTTCTCTG	129	5	187462	1
TAGTTGCTCC	130	7	317108	1
GTAAGAAATC	131	7	836202	1
CTTGGGCTAT	132	8	107992	1
AAATGGTGAT	133	11	558688	1
ATCATTGCGG	134	12	199358	1
CTGAACCTTA	135	12	283720	1
CCAGAAGGAG	136	13	652873	1
CCGGTTACTA	137	15	803663	1
CGATGAGAAG	138	15	1004369	1
AAACCGTCCC	139	16	199141	1
TCATTCATAC	140	2	164728	1
TATCTTTTTG	141	4	169784	1
TTAGAATAAT	142	4	603508	1
GTACGCTGTG	143	5	118089	1
TATATTAATT	144	6	64228	1

GTTCTTGCCT	145	7	939579	1
ATATAGCTGC	146	10	181144	1
CCAAAAA	147	11	91785	1
GAACTCCACA	148	11	94125	1
CCTTCACTGC	149	11	374172	1
CACATCATAA	150	11	625896	1
GAAGTATTGA	151	12	603999	1
TGCGCGTATA	152	13	206410	1
GGGTAGTACT	153	13	671730	1
TAGTTTTGTC	154	15	33475	1
CAATTCCTAC	155	1	172182	0.8
TTTGATTTGA	156	2	46431	0.8
GGCTCTGGTT	157	2	414510	0.8
CAGAAATAGC	158	2	565130	0.8
CTGTTATTTT	159	2	616054	0.8
CGAAGTCAAA	160	2	680605	0.8
CTCTAGATAA	161	3	171584	0.8
AGTCAAAATG	162	4	192750	0.8
GCGAGTTTAG	163	4	691301	0.8
GCTCCAATAG	164	4	1131020	0.8
TTTATTTGAG	165	4	1237501	0.8
GTTATATTGA	166	4	1401803	0.8
TGGGTTGAAG	167	5	251266	0.8
ATTTTATTTG	168	5	447729	0.8
ATCATAAAAA	169	5	548612	0.8
TTATATAAAA	170	6	223182	0.8
CTACTTCTGC	171	8	34653	0.8
ATAAGACAGT	172	10	227802	0.8
TTCATAAGTT	173	10	471894	0.8
TAAATCTGAG	174	11	145617	0.8
CTGGTAGAAA	175	11	151174	0.8
CACGTACACA	176	11	403208	0.8
CCAAGATCAA	177	11	425882	0.8
AGCTTGTTCC	178	12	234966	0.8
CACATTCGTT	179	12	759953	0.8
CTTACATATA	180	12	789781	0.8
TCTATAGCAA	181	13	228936	0.8
CCTTTCTGAA	182	13	297985	0.8
CCTTTAGAAT	183	13	777999	0.8
AATTAACACC	184	13	842122	0.8
GCGCAGGGGC	185	14	440984	0.8
TGTTTATAAA	186	14	661710	0.8
AAAAGTCATT	187	15	32081	0.8
TTCGTAAACT	188	15	680625	0.8
TTTTTGGAGT	189	15	888343	0.8
AGGCATCTTG	190	16	250284	0.8
AAATCAAAAC	191	16	453890	0.8
AATTGACGAA	192	16	560169	0.8
TTGATGATTT	193	16	582360	0.8
CCTGTTTTTG	194	16	643476	0.8
TTTTTAAAAA	195	1	101436	0.5

AAGTTTGATC	196	1	199848	0.5
AGCACCTATG	197	2	46913	0.5
TGATTTATCC	198	2	418946	0.5
ACTGCATCTG	199	2	680860	0.5
CAAGTTAGGA	200	2	744770	0.5
ATACCCAATT	201	3	29939	0.5
AACTTTGTAT	202	3	30056	0.5
GCGGCGGGTG	203	3	41645	0.5
AAAATTGTTC	204	3	57108	0.5
TCAAGTACTC	205	3	157855	0.5
AACTGTATGC	206	3	223882	0.5
CTATCGGCCA	207	3	278840	0.5
ACAAGCCCAA	208	3	289917	0.5
GTACAGGGCT	209	4	93873	0.5
AAGATCATCG	210	4	254851	0.5
GAATCCTGG	211	4	340891	0.5
GAACGAGAAG	212	4	371850	0.5
TTTTTAATAC	213	4	372058	0.5
TCTCCAGTTG	214	4	381712	0.5
AATACGTTAC	215	4	471791	0.5
ACGATTGGCT	216	4	509158	0.5
TGTTTATAAG	217	4	521709	0.5
CGTTTTCGTC	218	4	538839	0.5
TCGAACCTCT	219	4	578702	0.5
TCCACACACA	220	4	930972	0.5
CCGTGCGTGC	221	4	1324367	0.5
TTTCTTCAAC	222	5	116099	0.5
CCAAGTCTCG	223	5	159320	0.5
AGAGCGAATT	224	5	207517	0.5
TGTAGATTAT	225	5	280465	0.5
AAAAGTAGTT	226	5	286387	0.5
ACTTGGTATG	227	5	422942	0.5
TTAATGTTAT	228	5	544523	0.5
TACACGCGCG	229	5	544555	0.5
GGTCACTCCT	230	6	62983	0.5
AAGTGATGAA	231	6	76141	0.5
TTTATCTTGT	232	6	130327	0.5
AGTGATTGTT	233	6	256223	0.5
GCTTTGTTGT	234	7	72577	0.5
TCATTGATTC	235	7	110590	0.5
TTCACCGGAA	236	7	323655	0.5
ACTATTCTGT	237	7	423957	0.5
GGGCCAACCC	238	7	433787	0.5
AAAATATCTT	239	7	559397	0.5
TAGTAGTAAC	240	7	622201	0.5
AAGCGCACAA	241	7	735909	0.5
TCGCTGTTTT	242	7	800300	0.5
TGTATTTTTG	243	7	836202	0.5
CTAAACAAAG	244	7	836587	0.5
TAGGAAGAAA	245	7	905046	0.5
GGAAAAATTA	246	7	958839	0.5

TTTGATAGT	247	7	974754	0.5
CGTTTGTGTA	248	8	202655	0.5
AGAAAAAAC	249	8	386651	0.5
TAAAGTCCAG	250	8	518998	0.5
TAAGCAGATT	251	8	529129	0.5
ATGAGCATT	252	9	97114	0.5
AGGTGCAAAA	253	9	229077	0.5
TAACAAAGAG	254	10	628227	0.5
CAATTGGCAA	255	10	721781	0.5
ACTCCCTGTA	256	11	93528	0.5
CTCTATTGAT	257	11	144281	0.5
GCTTTCCTT	258	11	146665	0.5
ACCGCAAAGA	259	11	231872	0.5
CTTGTTCAAA	260	12	230972	0.5
AATGTGCTGT	261	12	320426	0.5
GCAGATAGCG	262	12	341324	0.5
TCTGACTTAG	263	12	368780	0.5
CCCGGATGTT	264	12	433912	0.5
GTAACGATTG	265	12	449917	0.5
GAATAACGAA	266	12	673851	0.5
ACTGCTATTT	267	12	712478	0.5
GTTCTCTAGC	268	12	712712	0.5
CATCACCATC	269	12	784710	0.5
TTGCACTTCT	270	12	806833	0.5
ACTGTTTATG	271	12	867350	0.5
TTGCTATATA	272	12	1017911	0.5
TACATTCTAA	273	13	95707	0.5
CTCTAGTTG	274	13	158970	0.5
ACGAACACTT	275	13	278341	0.5
TGCGCAAGTC	276	13	283795	0.5
TTTTTCTTAA	277	13	363037	0.5
CAAATGCATT	278	13	390802	0.5
CAAATTGTGT	279	13	395599	0.5
GCAATACTAT	280	13	826521	0.5
AGTGACGATG	281	14	60143	0.5
TACTGGTTTA	282	14	118854	0.5
GTTTGACCTA	283	14	335512	0.5
AGCGTTTGAT	284	14	478481	0.5
CTCTGTTGCG	285	14	728251	0.5
AAATTCAAAA	286	15	35952	0.5
TTTGCTTGGT	287	15	242742	0.5
AGTTTTCTTG	288	15	304813	0.5
TTTAAAGATA	289	15	331453	0.5
AAGGAGACAC	290	15	448624	0.5
CTATATATCA	291	15	544530	0.5
GATGGAATAG	292	15	571210	0.5
TCGAGTCGAA	293	15	758202	0.5
AAAAAAGAAA	294	15	882567	0.5
TTTCCAGAAT	295	15	969884	0.5
TGGACAATGT	296	15	970607	0.5
GGAATTAAGA	297	15	979894	0.5

WO 00/77214

PCT/US00/16223

ACTATATGTT	298	16	582230	0.5
GATATATCAT	299	16	589647	0.5
AGAATTGATT	300	16	744406	0.5
CACTGTCTCC	301	16	824649	0.5

Table 6, cont.

11	91621	91756	CCAAAAA	91785	1.1	580
11	91700	91826	CCAAAAA	91785	1.1	581
11	93604	93778	GAACCTCCACA	94125	1.1	582
11	93620	93731	GAACCTCCACA	94125	1.1	583
11	93653	93539	ACTCCCTGTA	93528	-0.5	584
11	93822	93732	ACTCCCTGTA	93528	-0.5	585
11	94057	93919	ACTCCCTGTA	93528	-0.5	586
11	94076	94229	GAACCTCCACA	94125	1.1	587
11	94111	94234	GAACCTCCACA	94125	1.1	588
11	94127	93998	ACTCCCTGTA	93528	-0.5	589
11	143708	143804	CTCTATTGAT	144281	0.5	590
11	143803	143929	CTCTATTGAT	144281	0.5	591
11	146079	145929	TAAATCTGAG	145617	-0.8	592
11	146115	146196	GCITTCCTTT	146665	0.5	593
11	146591	146756	GCITTCCTTT	146665	0.5	594
11	151395	151557	CTGGTAGAAA	151174	0.8	595
11	202891	202981	AACAAGATTG	203396	1.1	596
11	203065	203146	AACAAGATTG	203396	1.1	597
11	212738	212657	TCTGTGAAAC	212642	-1.9	598
11	212805	212706	TCTGTGAAAC	212642	-1.9	599
11	212887	212773	TCTGTGAAAC	212642	-1.9	600
11	213019	212944	TCTGTGAAAC	212642	-1.9	601
11	226029	226119	GAGGCAAAA	226288	0.5	602
11	231305	231446	ACCGCAAAGA	231872	0.5	603
11	316824	316716	TCTATTGTCA	316337	-4.8	604
11	374386	374269	CCTCACTGC	374172	-1.1	605
11	374400	374499	CTGTTTTGGG	374761	1.9	606
11	374465	374339	CCTCACTGC	374172	-1.1	607

References

Archer, J. E., Vega, L. R., and Solomon, F. (1995). Rbl2p, a yeast protein that binds to beta-tubulin and participates in microtubule function in vivo. *Cell* 82, 425-434.

5 Bajwa, W., Torchia, T. E., and Hopper, J. E. (1988). Yeast regulatory gene GAL3: carbon regulation; UASGal elements in common with GAL1, GAL2, GAL7, GAL10, GAL80, and MEL1; encoded protein strikingly similar to yeast and *Escherichia coli* galactokinases. *Mol Cell Biol* 8, 3439-3447.

10 Basrai, M. A., Kingsbury, J., Koshland, D., Spencer, F., and Hieter, P. (1996). Faithful chromosome transmission requires Spt4p, a putative regulator of chromatin structure in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16, 2838-2847.

Bishop, J. O., Morton, J. G., Rosbash, M., and Richardson, M. (1974). Three abundance classes in HeLa cell messenger RNA. *Nature* 250, 199-204.

15 Burkholder, A. C., and Hartwell, L. H. (1985). The yeast alpha-factor receptor: structural properties deduced from the sequence of the STE2 gene. *Nucleic Acids Res* 13, 8463-8475.

Chambers, A., Tsang, J. S., Stanway, C., Kingsman, A. J., and Kingsman, S. M. (1989). Transcriptional control of the *Saccharomyces cerevisiae* PGK gene by RAP1. *Mol Cell Biol* 9, 5516-5524.

20 Denis, C. L., Ferguson, J., and Young, E. T. (1983). mRNA levels for the fermentative alcohol dehydrogenase of *Saccharomyces cerevisiae* decrease upon growth on a nonfermentable carbon source. *J Biol Chem* 258, 1165-1171.

25 Dick, T., Surana, U., and Chia, W. (1996). Molecular and genetic characterization of SLC1, a putative *Saccharomyces cerevisiae* homolog of the metazoan cytoplasmic dynein light chain1. *Mol Gen Genet* 251, 38-43.

El-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., Lin, D., Mercer, W. E., Kinzler, K. W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. *Cell* 75, 817-825.

30 Elledge, S. J., and Davis, R. W. (1989). DNA damage induction of ribonucleotide reductase. *Mol Cell Biol* 9, 4932-4940.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H.,

Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S.G. (1996). Life with 6000 genes. *Science* 274, 546-567.

5 Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63, 751-762.

10 Hagen, D. C., McCaffrey, G., and Sprague, G. F., Jr. (1986). Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone α factor: gene sequence and implications for the structure of the presumed receptor. *Proc Natl Acad Sci U S A* 83, 1418-1422.

Herford, L. M., and Rosbash, M. (1977). Number and distribution of polyadenylated RNA sequences in yeast. *Cell* 10, 453-462.

15 Irriger, S., and Braus, G. H. (1994). Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 91, 257-261.

Iyer, V., and Struhl, K. (1996). Absolute mRNA levels and transcriptional initiation rates in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 93, 5208-5212.

20 Kurjan, J., and Herskowitz, I. (1982). Structure of a yeast pheromone gene (MF α): a putative α -factor precursor contains four tandem copies of mature α -factor. *Cell* 30, 933-943.

Leeds, P., Peltz, S. W., Jacobson, A., and Culbertson, M. R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* 5, 2303-2314.

25 Lewin, B. (1980). *Gene Expression* 2, (New York, New York: John Wiley and Sons), pp. 694-727.

McAlister, L., and Holland, M. J. (1982). Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. *J Biol Chem* 257, 7181-7188.

30 Michaelis, S., and Herskowitz, I. (1988). The α -factor pheromone of *Saccharomyces cerevisiae* is essential for mating. *Mol Cell Biol* 8, 1309-1318.

Mushegian, A. R., and Koonin, E. V. (1996). A minimal gene set for cellular life

derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* 93, 10268-10273.

Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., and Jordan, B. R. (1995). Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* 29, 207-216.

Nishizawa, M., Araki, R., and Teranishi, Y. (1989). Identification of an upstream activating sequence and an upstream repressible sequence of the pyruvate kinase gene of the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 9, 442-451.

Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K., and Gottschling, D. E. (1993). Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev* 7, 1133-1145.

Rose, M. D., Winston, F., and P. Hieter. (1990). *Methods in Yeast Genetics*. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 177.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.

Schmitt, H. D., Ciriacy, M., and Zimmermann, F. K. (1983). The synthesis of yeast pyruvate decarboxylase is regulated by large variations in the messenger RNA level. *Mol Gen Genet* 192, 247-252.

Sikorski, R. S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19-27.

Singh, A., Chen, E. Y., Lugovoy, J. M., Chang, C. N., Hitzeman, R. A., and Seeburg, P. H. (1983). *Saccharomyces cerevisiae* contains two discrete genes coding for the alpha-factor pheromone. *Nucleic Acids Res* 11, 4049-4063.

Smith, M. M., and Murray, K. (1983). Yeast H3 and H4 histone messenger RNAs are transcribed from two non-allelic gene sets. *J Mol Biol* 169, 641-661.

St John, T. P., and Davis, R. W. (1979). Isolation of galactose-inducible DNA sequences from *Saccharomyces cerevisiae* by differential plaque filter hybridization. *Cell* 16, 443-452.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.

Zaret, K. S., and Sherman, F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell* 28, 563-573.

CLAIMS

1. An isolated DNA molecule comprising a coding sequence of a yeast gene selected from the group of NORF genes comprising a SAGE tag as shown in SEQ ID NOS:67-811.

5 2. The isolated DNA molecule of claim 1 which is involved in cell cycle progression.

3. The isolated DNA molecule of claim 2 wherein expression of the NORF gene varies by at least 10% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

10 4. The isolated DNA molecule of claim 2 wherein expression of the NORF gene varies by at least 25% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

15 5. The isolated DNA molecule of claim 2 wherein expression of the NORF gene varies by at least 50% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

6. The isolated DNA molecule of claim 2 wherein expression of the NORF gene varies by at least 100% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

20 7. The isolated DNA molecule of claim 2 wherein expression of the NORF gene varies by a statistically significant difference (greater than 95% confidence level) between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

8. The isolated DNA molecule of claim 7 wherein the NORF gene is selected from the group consisting of NORF N^o 1, 2, 4, 5, 6, 17, 25, and 27.

25 9. The isolated DNA molecule of claim 2 wherein the NORF gene is not expressed in at least one phase of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

10. The isolated DNA molecule of claim 1 which is genomic.

11. The isolated DNA molecule of claim 1 which is cDNA.

30 12. A method of using NORF genes to affect the cell cycle, comprising the

step of:

administering to a cell an isolated DNA molecule comprising a coding sequence of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M.

13. The method of claim 12 wherein the cell is a yeast cell.

14. The method of claim 12 wherein the cell is a fungal cell.

15. The method of claim 12 wherein the cell is a mammalian cell.

16. The method of claim 12 wherein the NORF gene is selected from the group consisting of NORF N^o 1, 2, 4, 5, 6, 17, 25, and 27.

17. A method for screening candidate antifungal drugs, comprising the steps of:

contacting a test substance with a yeast cell;

monitoring expression of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M, wherein a test substance which modifies the expression of the yeast gene is a candidate antifungal drug.

18. The method of claim 17 wherein the NORF gene is selected from the group consisting of NORF N^o 1, 2, 4, 5, 6, 17, 25, and 27.

19. A method for identifying human genes which are involved in cell cycle progression, comprising the steps of:

contacting human DNA with a probe which comprises at least 10 contiguous nucleotides of a NORF gene whose expression varies by at least 10% between any two phases of the cell cycle selected from the group consisting of log phase, S phase, and G2/M phase, wherein a human DNA sequence which hybridizes to the probe is identified as a sequence of a candidate human gene which is involved in cell cycle progression.

20. The method of claim 19 wherein the NORF gene is selected from the group consisting of NORF N^o 1, 2, 4, 5, 6, 17, 25, and 27.

21. A probe comprising at least 14 contiguous nucleotides of a NORF gene comprising a SAGE tag as shown in SEQ ID NOS:67-811.

22. The probe of claim 21 wherein expression of the NORF gene varies by at least 10% between any two phases of a cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

5 23. The probe of claim 22 wherein expression of the NORF gene varies by at least 25% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

24. The probe of claim 22 wherein expression of the NORF gene varies by at least 50% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

10 25. The probe of claim 22 wherein expression of the NORF gene varies by at least 100% between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

15 26. The probe of claim 22 wherein the NORF gene is not expressed in at least one phase of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

27. The probe of claim 22 wherein expression of the NORF gene varies by a statistically significant difference (greater than 95% confidence level) between any two phases of the cell cycle selected from the group consisting of: log phase, S phase, and G2/M.

20 28. The probe of claim 22 wherein the gene is selected from the group consisting of NORF N^o 1, 2, 4, 5, 6, 17, 25, and 27.

29. The method of claim 17 wherein said step of monitoring expression is performed using nucleic acid molecules which are immobilized on a solid support.

25 30. The method of claim 29 wherein the nucleic acid molecules are in on array.

31. The method of claim 19 wherein a probe which comprises a portion of the NORF gene is in an array on a solid support.

30 32. An array of probes on a solid support wherein at least one probe comprises at least 14 contiguous nucleotides of a NORF gene comprising a SAGE tag as shown in SEQ ID NOS:67-811.

33. The array of claim 32 wherein the at least one NORF gene is involved

in cell cycle progression.

34. The array of claim 32 wherein the NORF gene is selected from the group consisting of NORF No. 1, 2, 4, 5, 6, 17, 25, and 27.

5 35. The array of claim 32 which comprises at least 100 probes of distinct sequence.

36. The array of claim 32 which comprises at least 500 probes of distinct sequence.

37. The array of claim 32 which comprises at least 1,000 probes of distinct sequence.

10 38. A method of identifying a candidate drug as a member of a class of drugs having a characteristic effect on gene expression in a yeast cell, comprising the steps of:

contacting a yeast cell with a candidate drug; and

15 monitoring expression in the yeast cell of at least one NORF gene whose expression is affected by the class of drugs, wherein detection of a difference in expression of the at least one NORF gene in the yeast cell relative to expression in the absence of the candidate drug identifies the candidate drug as a member of the class of drugs.

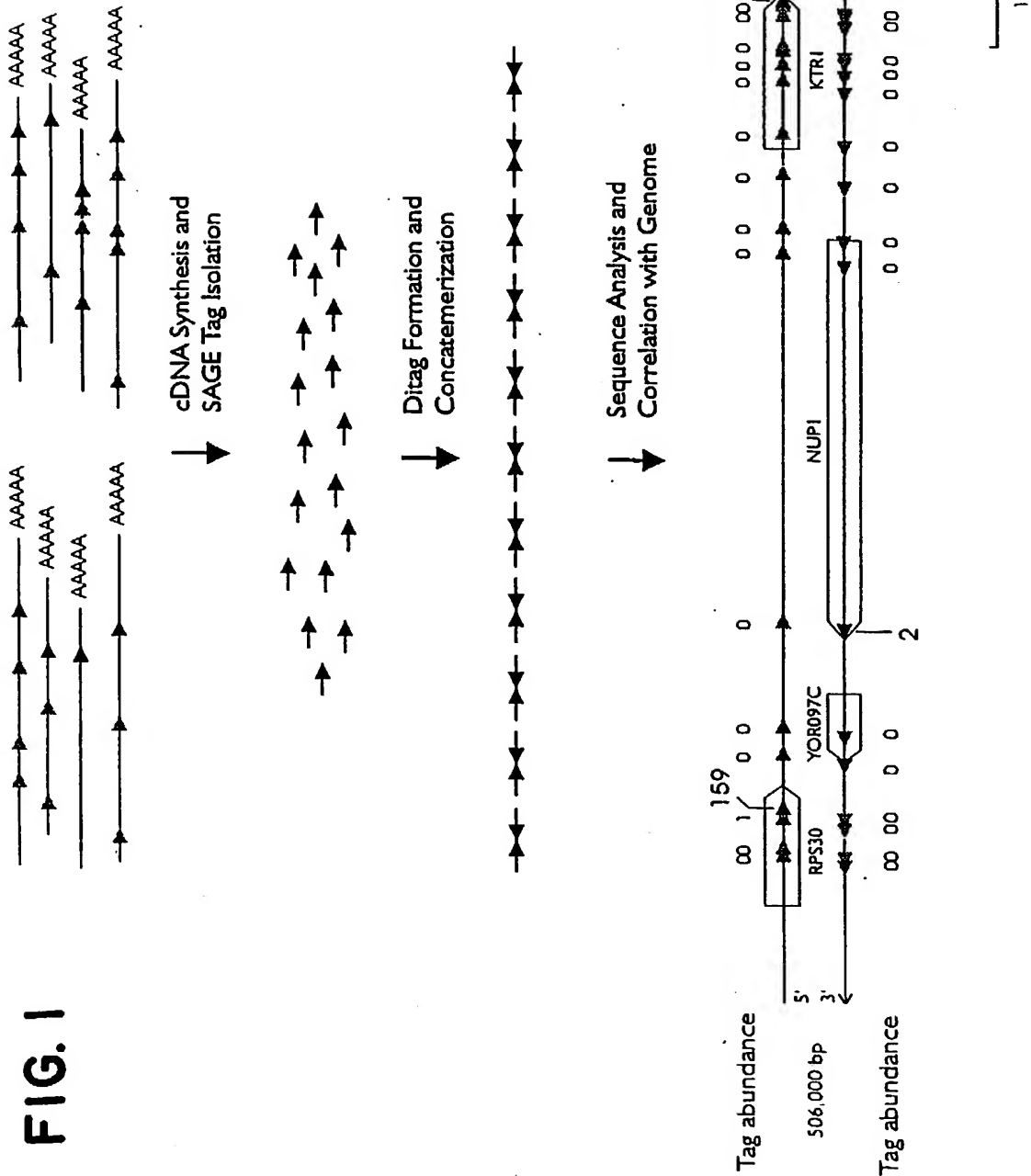
20 39. The method of claim 38 wherein the step of monitoring expression is performed using nucleic acid molecules which are immobilized on a solid support.

40. The method of claim 39 wherein the nucleic acid molecules are in an array.

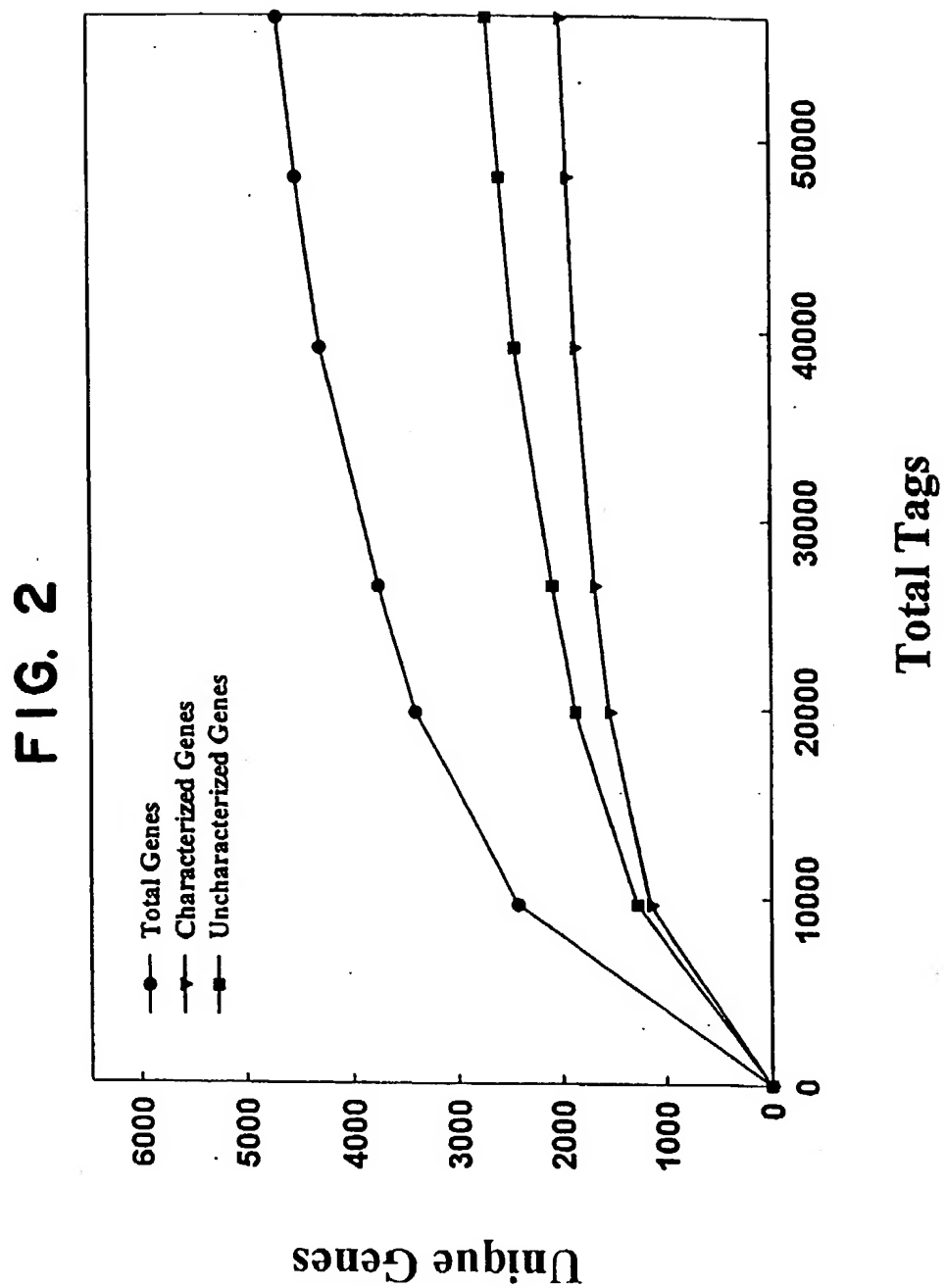
41. The method of claim 38 wherein expression of two or more NORF genes is monitored.

25 42. The probe of claim 21 which is immobilized on a solid support.

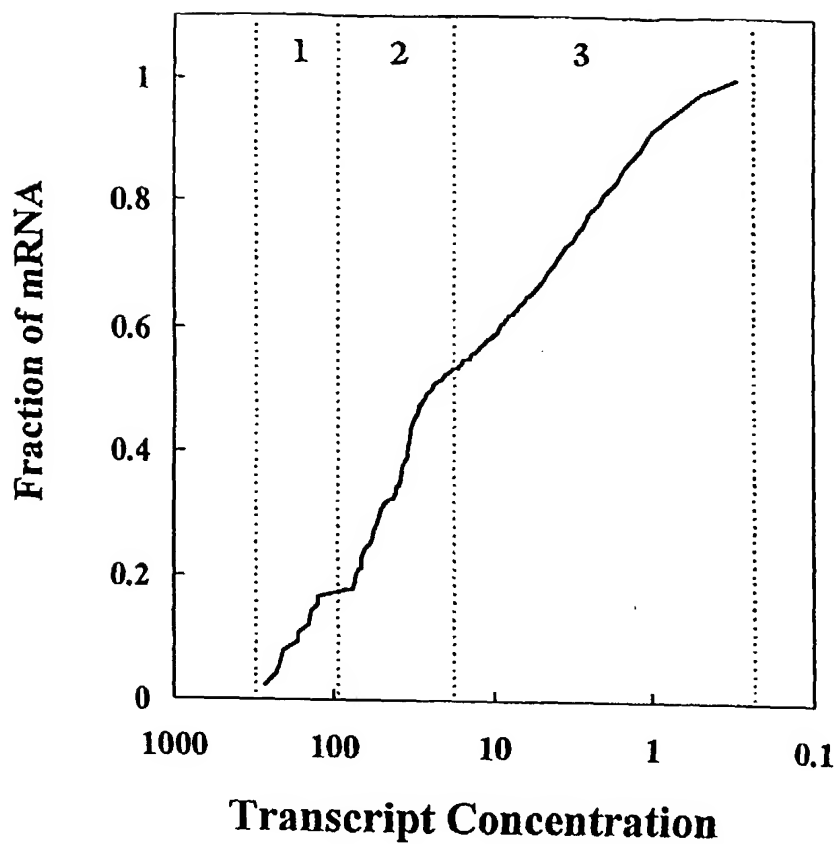
FIG. 1



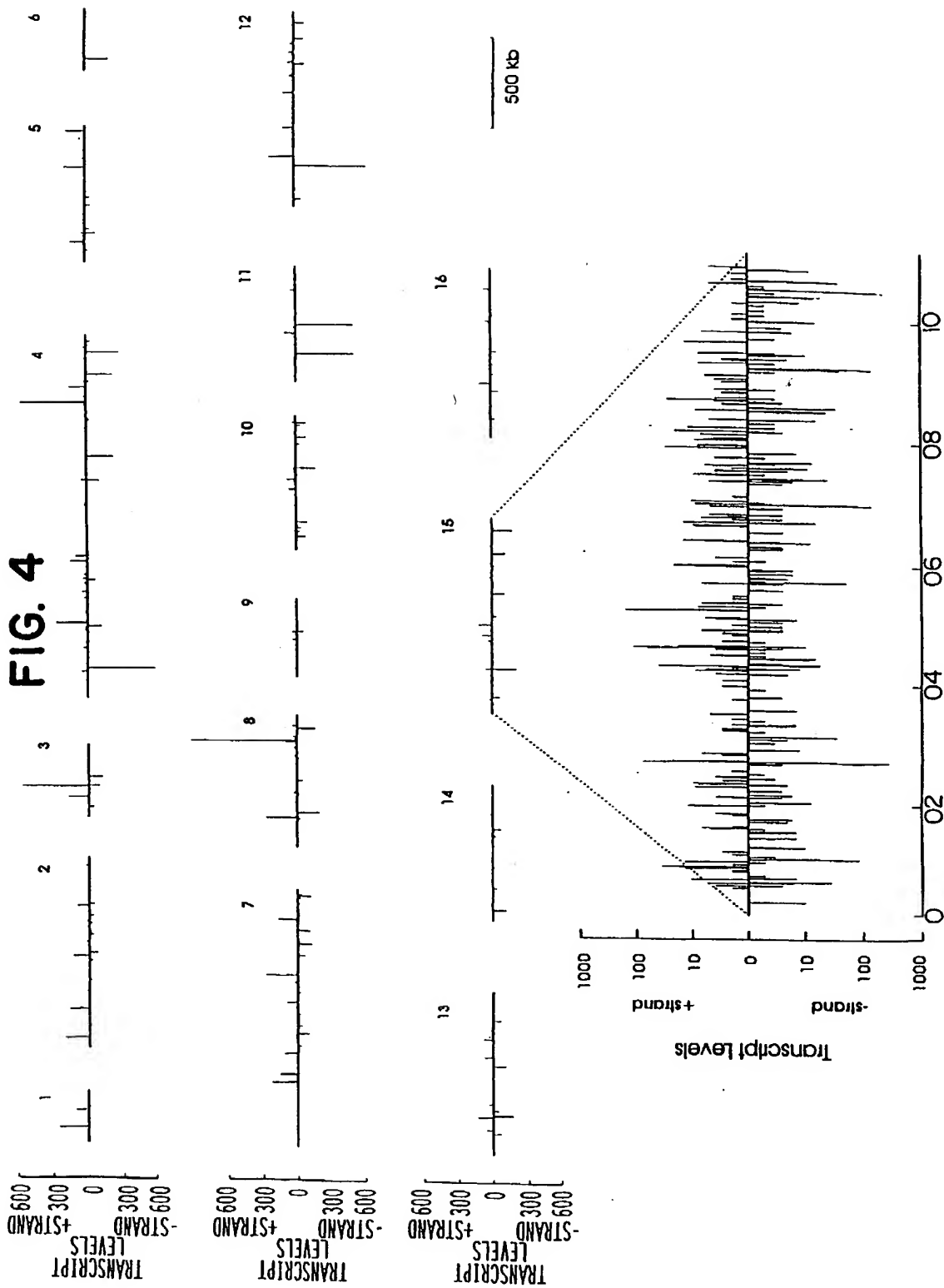
2/5



3/5

FIG. 3A**FIG. 3B**

Component	Virtual Rot (SAGE)		Rot (Reassociation)	
	%mRNA	Copies/cell	%mRNA	Copies/cell
1	17	180	23	200
2	38	40	51	30
3	45	2.5	26	1.5



5/5

FIG. 5